Nucleic Acids Research, 1993, Vol. 21, No. 9 2201–2208

Crystal structure of a Z-DNA hexamer d(CGCICG) at 1.7 Å resolution: inosine·cytidine base-pairing, and comparison with other Z-DNA structures

Vinod D.Kumar and Irene T.Weber
Department of Pharmacology, Bluemle Life Sciences Building, Jefferson Cancer Institute, Thomas Jefferson University, 233 S. 10th Street, Philadelphia, PA 19107, USA

Received December 7, 1992; Revised and Accepted March 8, 1993

ABSTRACT

The crystal structure of the deoxyhexamer, d(CGCICG), has been determined and refined to a resolution of 1.7Å. The DNA hexamer crystallises in space group $P2_12_12_1$ with unit cell dimensions of a = 18.412 ± .017 Å, b = 30.485 ± .036Å, and c = 43.318 ± .024 Å. The structure has been solved by rotation and translation searches and refined to an R-factor of 0.148 using 2678 unique reflections greater than 1.0 σ (F) between 10.0 – 1.7 Å resolution. Although the crystal parameters are similar to several previously reported Z-DNA hexamers, this inosine containing Z-DNA differs in the relative orientation, position, andcrystal packing interactions compared to d(CGCGCG) DNA. Many of these differences in the inosine form of Z-DNA can be explained by crystal packing interactions, which are responsible for distortions of the duplex at different locations. The most noteworthy features of the inosine form of Z-DNA as a result of such distortions are: (1) sugar puckers for the inosines are of C4′-*exo* type, (2) all phosphates have the Z_I conformation, and (3) narrower minor grove and compression along the helical axis compared to d(CGCGCG) DNA. In addition, the substitution of guanosine by inosine appears to have resulted in Watson – Crick type base-pairing between inosine and cytidine with a potential bifurcated hydrogen bond between inosine N1 and cytidine N3 (2.9 Å) and O2 (3.3 – 3.Å).

INTRODUCTION

Inosine is a purine nucleoside whose neutral base forms stable base pairs with all four conventional bases, and the strength of the base pairing is approximately equal in each case (1). Inosine occurs naturally in the wobble position of the anticodon of some t-RNAs, where it appears to pair with adenosine, thymidine or cytidine without destabilising the double helix. Poly(rI) and poly(dI) form stable helices with poly(rC) and poly(dC) (2), and serve as templates for the incorporation of cytosine into products of DNA and RNA polymerases (3). Oligonucleotides containing inosines have been used extensively as probes to screen human cDNA or genomic DNA libraries (4). These probes are generally used to reduce the degeneracy of pools of oligonucleotides, which has the advantage of reducing the possibility of mismatches and allowing the use of more stringent hybridization conditions. Despite this widespread use of inosine, it is becoming clear that inosine-containing base-pairs occur in a variety of configurations. The non-Watson Crick base pairs of *Ianti*:A*syn* mismatches (5) and I:T wobble base-pairs (6) have been observed as well as I:C Watson Crick base-pairs (7;8).

Inosine has a purine base that resembles guanine but lacks the 2-amino group. Base pairing between I and C is possible by forming two hydrogen bond interactions, instead of the three that occur in C:G base pairs. Oligomers of alternating CG sequences have been shown to form left-handed helices in solution at high ionic strength (9) and in the crystalline state (10;11). Previous studies using fiber diffraction have shown that poly dI.poly dC formed B-DNA double helices (12). In contrast, circular dichroism spectra showed an unusual negative band at high wavelength for poly d(I-C).poly d(I-C) which indicated a left-handed helical structure (13). More recently, Vorlickova and Sagi (14) have studied the effects of salt concentration on the conformation of poly d(I-C) using circular dichroism spectroscopy, and have observed A-, B-, and Z-DNA conformations, as well as an unusual conformation at low salt. We have previously reported the crystal structure of Z-DNA octamer d(CGCICICG) which contains two I:C base pairs (7). However, the hydrogen bond geometry for I:C base-pairs was indistinguishable from that of G:C base-pairs due to disorder in the crystal lattice. We describe here the high resolution crystal structure of an ordered Z-conformation hexanucleotide DNA containing I:C base pairs, which provides a more complete description of the (inosine-cytidine) base pairing and compare the structure with other published d(CGCGCG) Z-DNA structures (10;15). These Z-DNA structures are referred to as mixed magnesium/spermine and magnesium form respectively.

MATERIAL AND METHODS

Synthesis and crystallization

The oligonucleotide, d(CGCICG), was synthesized by solid phase phosphoramidite chemistry on an automated Applied Biosystems Synthesizer, purified by ion-exchange liquid chromatography and reverse-phase high pressure liquid chromatography, and assayed by polyacrylamide gel electrophoresis. Crystals were grown from a solution containing 1.5 microliter of 2mM Oligonucleotide, and 1.5 microliter of the reservoir solution containing 1mM Spermine, 14 mM $MgCl_2$, and 17% 2-Methyl,2,4,Pentanediol, at room temperature by micro-vapor diffusion in the hanging drop method. After one week, the concentration of 2-Methyl-2,4-Pentanediol in the reservoir was increased to 20%. Large single crystals began to appear the following day. The space group and the unit cell dimensions of the crystals, were determined by precession photographs. The unit cell dimensions of the crystals were $a=18.412\pm.017$ Å, $b=30.485\pm.036$Å, and $c=43.318\pm.024$ Å which is similar to the crystals of the d(CGCGCG) hexamer.

Data collection and reduction

A single crystal with approximate dimensions $0.42\times0.25\times0.15$ mm sealed in a quartz capillary was used to collect data to 1.7 Å resolution on a R-AXIS IIC imaging plate mounted on a Rigaku RU-200 rotating anode generator, operated at 50kV and 100 mA. The imaging plate was set at a distance of 70 mm from the center of the crystal. The cell constants $a=18.412\pm.017$ Å, $b=30.485\pm.036$Å, and $c=43.318\pm.024$ Å and the space group $P2_12_12_1$ are determined from the reflections recorded from three still photographs taken at 25° intervals at 15 minute exposure time. A total of 11376 observations were measured from twenty oscillation frames at 5° intervals and 15 minutes exposure times. The data reduction and scaling were performed with the program package PROCESS (16). The unweighted merging R value was 6.4% with 2702 independent reflections greater than 1.0 s(I) to a resolution of 1.7 Å, which represents 92% of the total possible reflections at that resolution.

Determination of structure

Molecular replacement. The overall similarity in the space group and unit cell dimensions with the other Z-DNA hexamers suggested an isomorphous structure. The initial R-factor obtained with a structure factor calculation using DNA coordinates from the mixed magnesium/spermine form of d(CGCGCG) DNA was 54% for 8−3.0 Å diffraction data. This value obviously is much higher than expected for two crystal structures with analogous orientations and positions of the DNA duplexes. The correct orientation of the DNA duplexes was determined by using the molecular replacement method incorporated into the XPLOR computer program (17).

(1) *Rotation search.* The rotation search was carried out using the Patterson search in XPLOR. The probe Patterson maps used structure factors calculated from the mixed magnesium/spermine form d(CGCGCG) duplex coordinates in an orthorhombic cell with a=36.0 Å, b=60.0 Å, and c=88.0 Å edges. One thousand of the highest Patterson vectors in the range 8−3.0 Å were selected and rotated using pseudoorthogonal Eulerian angles (θ_+, θ_2, θ_-) as defined by Lattman (18). The rotation search was restricted to the asymmetric unit $\theta_- = 0-180°$, $\theta_2 = 0-90°$, and $\theta_+ = 0-360°$ for $P2_12_12_1$ space group.

(2) *Patterson correlation refinement.* The next step was to carry out Patterson correlation (PC) refinement of the highest peaks of the rotation function. This procedure was carried out by minimization against a target function defined by Brunger (19) and implemented in XPLOR. From the initial rotation search the top 100 peaks were chosen for PC refinement. However after Patterson correlation refinement four strong peaks emerged which were different from the highest peak observed after the initial rotation search. These four peaks correspond to approximately the same orientation. It is clear that these four peaks were initially away from the correct orientation but converged to it during the PC refinement.

(3) *Translation function.* The translation search employed the standard linear correlation coefficient between the normalized observed structure factors and normalized calculated structure factors (19). The search was conducted using 8−3.0 Å data in the range x= 0−0.5, y = 0−0.5, and z = 0−0.5, with the sampling interval 0.022 of the unit cell length. Only a single position emerged at x = 0.386, y = 0.023, and z = 0.273 with a correlation coefficient of 0.6585. The initial R factor for the d(CGCGCG) coordinates in the determined molecular orientation and position was 0.42 for the 10.0−3.0 Å resolution data.

Structure refinement

The molecular replacement solution was initially refined by a 6-dimensional rigid body method to an R factor of 0.37 at 3.0 Å resolution, using XPLOR. At this stage, Fourier sum (2Fo−Fc) and difference maps (Fo−Fc) were displayed on an Evans & Sutherland ESV10 graphics terminal with the program FRODO (20). The electron density was nearly continuous around the DNA backbone and bases. The structure was initially refined in XPLOR until the R factor converged to 0.23 for 10−3.0 Å data. At this stage, refinement was continued according to the Hendrickson and Konnert (21) restrained least-squares refinement procedure with NUCLSQ (22). The refinement was extended to 1.7 Å in several stages. At each stage a series of Fourier sum and difference maps were calculated and displayed on the graphics terminal. Manual intervention was required during each of these stages to adjust the model to fit the electron density. After refinement to 2.2 Å, a series of difference Fourier maps were calculated at each stage and a peak search program was used to locate water molecules. The criteria for selection of solvent

Table I. Refinement parameters

Total number of non-hydrogen atoms	310		
DNA non-hydrogen atoms	240		
water non-hydrogen atoms	70		
Resolution range(Å)	10 - 1.7		
Reflections used $F_{obs} > 1\sigma(F_{obs})$	2678		
Final *R*-factor	0.148		
Weights	$w=\sigma_F^{-2}$		
with	$\sigma_F = (16.5) + (-90.0)*(s\text{-}1/6)$		
Distance restraints[a]			
Sugar/base bond distances	0.024	(0.020)	Å
Sugar/base angle distances	0.037	(0.030)	Å
Phosphate bond distances	0.050	(0.030)	Å
Phosphate bond angle distances	0.042	(0.030)	Å
Plane restraints[a]	0.051	(0.030)	Å
Chiral center restraints[a]	0.136	(0.100)	Å^3
Non-bonded restraints[a]			
single-torsion contact	0.056	(0.250)	Å
multiple-torsion contact	0.416	(0.250)	Å
B_{iso} restraints[a]			
Sugar-base bonds	3.427	(5.000)	Å^2
Sugar-base angles	4.006	(7.500)	Å^2
Phosphate bonds	5.772	(7.500)	Å^2
Phosphate bond angles	5.694	(7.500)	Å^2

[a]r.m.s. deviations from ideality (target restraints in parentheses)

molecules were: 1) a spherical peak in the difference Fourier map greater than 2.5 standard deviations in height; 2) potential hydrogen-bonding partners within 2.2 to 3.4 Å; and an acceptable thermal parameter of less than 60 $Å^2$ obtained after subsequent refinement. All solvent peaks were represented as water molecules, since it was not possible to specifically identify sodium, magnesium or spermine ions. Refinement restraints were applied to covalent bond lengths and bond angles, chirality of deoxyribose atoms, planarity of base atoms, nonbonded repulsive contacts and hydrogen-bonded contacts. The deoxyribose rings in the structure were fit to the electron density and allowed to refine without any restraints. Soft restraints were imposed on the individual isotropic temperature factors. The refined d(CGCICG) structure including 70 solvent water molecules has an R factor of 0.148 when all observed X-ray data (2678 reflections) between 10−1.7 Å with I> 1.0 σ(I) are included. Table I gives a list of the refinement parameters.

RESULTS AND DISCUSSION

Helix structure

The asymmetric unit consists of two chemically equivalent self-complementary hexanucleotide strands forming an anti-parallel duplex of the Z-DNA type. Thenucleosides are labeled C1 through G6 in the 5′ to 3′ direction in strand 1, and C7 through G12 on strand 2. The nucleosides G(4) and G(10) in the parent hexamer d(CGCGCG) are replaced by I in this structure. The strands from each of these molecules form a continuous double helix with its chemically equivalent self-complementary strand along the c-axis. The d(CGCICG) hexamer is rotated by 14.79° around the helical axis and translated by 16.89 Å in the direction of the helical axis compared to the mixed magnesium/spermine form d(CGCGCG) DNA structure (10). After simple superposition of all the atoms of d(CGCICG) and d(CGCGCG) DNA structures, the total root mean square (RMS) deviation of atomic positions for all atoms was between 0.60 and 0.63 Å, with the average increasing from the bases to the sugars and phosphates, respectively. The overall conformation of the inosine-substituted d(CGCICG) hexamer appears to be similar to that of the other known Z-DNA hexamer structures. However an overall RMS deviation reveals very little information on the distinct local conformational alternations of the DNA in different crystal forms. These differences are clearly revealed by superimposing the terminal base pairs (C1-G12) of d(CGCICG) DNA and the mixed magnesium/spermine form of the parent hexamer d(CGCGCG) structure (Figure 1). It is evident that the d(CGCICG) duplex is shorter than the mixed magnesium/spermine form. The average helical rise in the d(CGCICG) DNA is 3.66 Å compared to 3.75 Å and 3.83 Å for the magnesium form of d(CGCGCG) and mixed magnesium/spermine form of d(CGCGCG) DNA respectively (Table II). The d(CGCICG) duplex is thus shorter by 0.42 Å and 0.81 Å compared to the magnesium form and the mixed magnesium/spermine form respectively. Based on an average rise of 3.66 Å, a nominal complete turn of a Z-DNA helix (12 base pairs per turn) for the d(CGCICG) DNA would be 43.9 Å compared to 45.0 and 45.96 Å for the magnesium and magnesium/spermine form DNA respectively. The d(CGCICG) hexamer is however very similar in size to the pure spermine form d(CGCGCG) DNA which shows an average helical rise of 3.67 Å for a dinucleotide repeat, and a height of 44.04 Å for one complete turn of the helix.

Although d(CGCICG) DNA adopts a double-helical structure with a helical twist of −60° for one dinucleotide repeat similar to other Z-DNA hexamers, there is a major difference in the twist angles for the CpG(I) or G(I)pC steps (Table II). For the d(CGCGCG) molecule the average twist angle for the CpG pair is −8°, compared to a significantly large rotations of −51° for the GpC base pair. However, for the d(CGCICG) DNA the CpG(I) steps appear to be slightly overwound (−12°) while the G(I)pC steps are unwound (−48°) compared to d(CGCGCG) DNA. The net effect of these changes in the twist angle is to bring the the cytidine groups closer to the guanosine (inosine) residues (11).

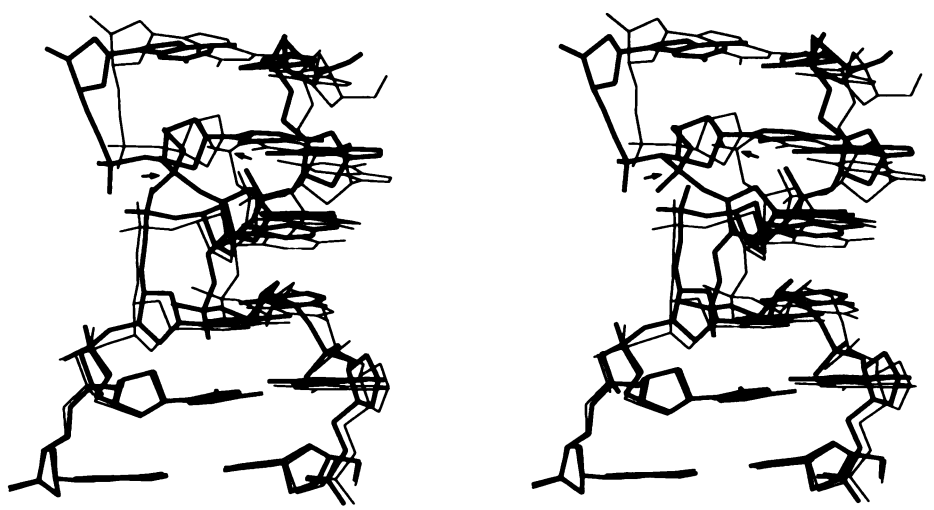

Figure 1. Comparison of the DNA conformations of the inosine-form d(CGCICG) DNA (thick lines) and the mixed magnesium/spermine form d(CGCGCG) DNA (thin lines). To emphasize the conformational deviations between the two structures, only the terminal base-pairs C1:G12 of the two structures were superimposed. Also indicated by arrows are the phosphate conformations for the central I4pC5 and I10pC11 steps. The mixed magnesium/spermine form d(CGCGCG) DNA reveals two separate phosphate orientations: The phosphate C5 on strand 1 is rotated away from the groove in Z_{II} conformation and the phosphate between G10 and C11 is rotated towards the groove in the helix in Z_I conformation. However, in d(CGCICG) DNA all phosphates are in the Z_I conformation.

Table II. Helical parameters of d(CGCICG)[a], mixed magnesium/spermine form d(CGCGCG)[b] and magnesium form d(CGCGCG)[c] DNA Structures

Sequence	tip	inclination	roll	tilt	propeller twist	buckle	twist	rise	X-disp
C1-G12	3.96	-4.56	-4.59	1.13	-1.50	**-5.30**	**11.84**	3.62	3.81
C1-G12	7.22	-8.22	-4.07	-1.55	-0.50	**-0.53**	**7.53**	3.94	2.83
C1-G12	6.57	-5.57	-4.62	-1.42	0.20	**-2.50**	**7.35**	3.81	3.02
G2-C11	-0.62	-5.71	**4.74**	-0.54	1.74	**8.71**	49.19	3.81	3.31
G2-C11	3.20	-6.73	**-0.10**	-2.84	-3.39	**6.24**	49.81	3.90	2.26
G2-C11	1.98	-4.18	**-0.77**	-1.48	-3.89	**4.43**	50.40	3.83	2.63
C3-**I10**	4.10	-5.15	-5.81	-1.69	-4.96	**-8.65**	**11.92**	3.60	2.06
C3-G10	3.12	-3.89	-4.35	0.16	-6.01	**-4.29**	**8.03**	3.57	1.36
C3-G10	1.22	-2.70	-3.53	1.01	-1.76	**-4.46**	**7.73**	3.65	1.63
I4-C9	-1.69	-3.47	**4.39**	-0.57	1.02	**8.28**	48.15	3.65	2.30
G4-C9	-1.22	-4.05	**1.67**	0.21	-0.11	**8.17**	51.96	3.65	1.32
G4-C9	-2.31	-3.72	**1.91**	0.68	-0.13	**4.53**	51.83	3.61	1.58
C5-G8	2.70	-2.90	**3.38**	0.04	-1.87	**-3.56**	12.27	3.50	3.39
C5-G8	0.44	-4.26	**0.22**	2.31	-0.01	**0.18**	9.63	4.06	2.65
C5-G8	-0.40	-4.40	**-1.23**	0.94	-0.67	**-3.34**	10.70	3.84	2.67
G6-C7	**6.07**	-2.92	-	-	1.99	**0.69**	-	-	3.35
G6-C7	**1.10**	-6.57	-	-	4.75	**-5.38**	-	-	2.94
G6-C7	**-1.63**	-5.34	-	-	2.33	**0.08**	-	-	3.09
mean	2.99	-4.12	0.42	-0.33	-0.60	0.03	26.67	**3.66**	**3.04**
mean	2.31	-5.62	-1.24	-0.34	-0.88	0.73	25.39	**3.83**	**2.23**
mean	0.90	-4.32	-1.65	-0.05	-0.65	-0.21	25.60	**3.75**	**2.44**

Helical parameters were calculated with program NEWHELIX91 distributed by R.E.Dickerson. [a]Parameters for d(CGCICG) DNA are shown on the first line. [b]Parameters for mixed magnesium/spermine form d(CGCGCG) DNA (Wang et al. 1979) are given on the second line. [c]Parameters for magnesium form d(CGCGCG) DNA (Gessener et al 1989) are given on the third line. Notable differences are shown in bold.

I:C base-pair geometry

Figures 2A and 2B show a 2(Fo−Fc) Fourier electron density sum map calculated at 1.7 Å resolution around the C(3)−I(10) and I(4)−C(9) base pairs respectively. The base pairing between inosine and cytidine is of the Watson−Crick type in an anti-parallel helix. The angles between the glycosyl bond and the C1′...C1′ vector for the base-pairs are well within the narrow range of 52° to 62° observed for Watson−Crick base-pairs (Kennard 1988), suggesting no wobble base pairing between inosine and cytidine. However, a close inspection of the two I:C base-pairs showed that in addition to the hydrogen bonds formed between N1 of I4 and N3 of C9 (2.82 Å) and O6 of I4 with N4 of C9 (2.94 Å), N1 of I4 forms a second slightly longer hydrogen bond with O2 of C9 (3.34 Å). A similar hydrogen bonding scheme is observed with the second I:C base-pair, where N1 of I10 forms a hydrogen bond with O2 of C3 (3.38 Å). In the mixed magnesium/spermine form of d(CGCGCG) DNA the equivalent distances are longer, 3.58 Å and 3.69 Å respectively. These observations lead us to suggest the amino group N1 of inosine (I4 and I10) appears to participate in bifurcated hydrogen bonding interactions.

In contrast to the bifurcated hydrogen bonding observed in I:C base-pairs in d(CGCICG) structure, wobble base-pairing was observed between I:T in an A-DNA octanucleotide structure d(GGIGCTCC) (6), while non-Watson−Crick I (*anti*):A (*syn*) mismatched base-pairs were observed in the crystal structure of d(CGCIAATTAGCG) (5). Recently it has been shown that in the B-DNA, d(CGCIAATTCGCG) dodecamer (8), I:C base-pairs form Watson−Crick type hydrogen bonds, indicating that inosine can adopt a variety of base-pair configurations similar to those of guanosine.

Table IIIa. Sugar-phosphate backbone, glycoside torsion angles, and pseudorotation phase angles (P) of d(CGCICG)[a], mixed magnesium/spermine form d(CGCGCG)[b] and magnesium form d(CGCGCG)[c] DNA Structures

Sequence	α	β	γ	δ	ε	ζ	χ	P
Strand 1								
C1	-	-	**66**	**130**	272	83	204	148
C1	-	-	**47**	**144**	267	81	208	154
C1	-	-	**52**	**145**	265	79	210	154
G2	55	191	178	94	**254**	291	66	27
G2	63	186	174	94	**240**	294	57	30
G2	61	188	178	91	**240**	295	60	40
C3	206	**216**	**64**	140	269	85	197	150
C3	216	**234**	**50**	152	258	76	202	148
C3	212	**239**	**50**	148	260	81	210	145
I4	65	192	183	92	**227***	**326***	56	**59†**
G4	70	188	177	95	**181***	**65***	52	**30†**
G4	64	186	179	92	**181***	**69***	59	**26†**
C5	**180**	**222**	56	131	266	69	204	**132†**
C5	**169**	**167**	43	142	267	74	215	**155†**
C5	**166**	**160**	48	142	260	80	208	**151†**
G6	**96**	180	174	143	-	-	71	161
G6	**74**	178	180	149	-	-	79	162
G6	**76**	175	182	149	-	-	78	170

Table IIIb. Sugar-phosphate backbone, glycosidic torsion angles, and pseudorotation phase angles (P) of d(CGCICG)[a], mixed magnesium/spermine form d(CGCGCG)[b] and magnesium form d(CGCGCG)[c] DNA Structures

Sequence	α	β	γ	δ	ε	ζ	χ	P
Strand 2								
C7	-	-	**103**	141	257	**90**	210	161
C7	-	-	**53**	147	270	**78**	218	154
C7	-	-	**55**	139	268	**74**	209	157
G8	61	184	180	99	**264***	**294***	**51**	22
G8	61	187	175	95	**244***	**286***	**70**	29
G8	67	189	172	101	**236***	**335***	**61**	36
C9	203	**215**	56	139	265	79	215	158
C9	220	**225**	55	149	262	80	200	154
C9	198	**196**	55	140	268	74	206	153
I10	62	185	182	96	**219***	**313***	61	**57†**
G10	64	179	179	103	**248***	**290***	65	**18†**
G10	64	185	179	96	**244***	**290***	62	**35†**
C11	**193**	**223**	60	135	273	63	194	147
C11	**212**	**236**	50	143	262	72	204	147
C11	**210**	**241**	56	142	259	70	203	151
G12	89	182	185	152	-	-	73	166
G12	74	184	186	149	-	-	79	167
G12	84	183	183	149	-	-	72	162

Glycosidic torsion angles and pseudorotation phase angles were calculated with program NEWHELIX91 distributed by R.E.Dickerson. [a]Parameters for d(CGCICG) DNA are shown on the first line. [b]Parameters for mixed magnesium/spermine form d(CGCGCG) DNA (Wang et al. 1979) are given on the second line. [c]Parameters for magnesium form d(CGCGCG) DNA (Gessener et al 1989) are given on the third line. *show differences in backbone torsion angles ε and ζ due to Z_{II} conformation of P5 and a partial Z_{II} conformation of P9 and P11 in both the mixed magnesium/spermine and magnesium form d(CGCGCG) DNA. †show differences in pseudorotation phase angles. Other notable differences are shown in bold.

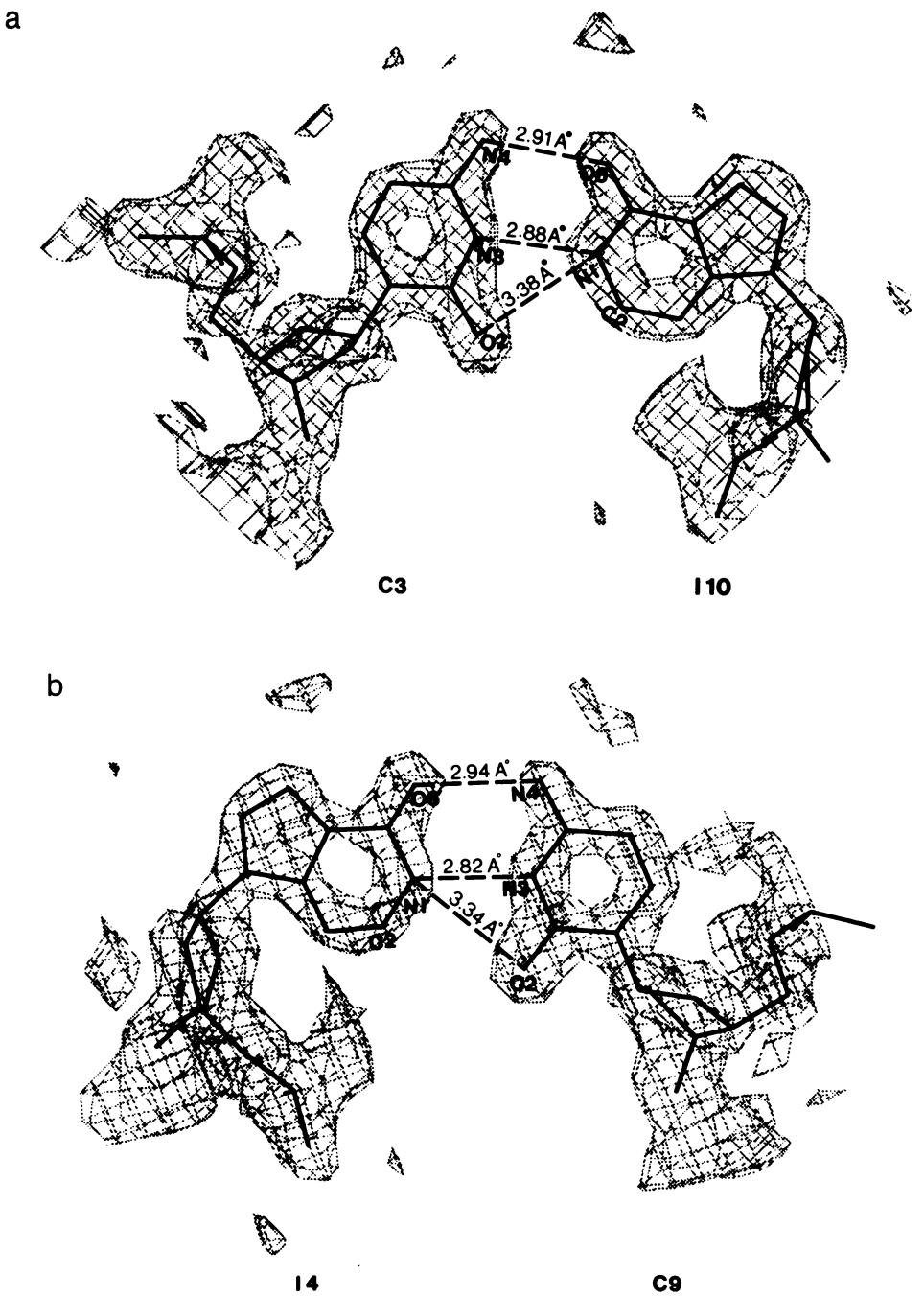

Figure 2. A C:I base pair super-imposed on a section of a 2Fo−Fc electron density map calculated after refinement of the structure to 1.7 Å resolution. The electron density is contoured at a level of 1.0 σ and is shown in dashed lines, while the atomic coordinates are indicated by continuous lines. a. C3:.I10; b. I4:C9

Base stacking

Table II compares various helical parameters for the d(CGCICG) DNA with the mixed magnesium/spermine and magnesium forms of d(CGCGCG) DNA based on the IUPAC-IUB nomenclature (1989). As indicated earlier, although the backbone is minimally perturbed by substituting guanosine with inosine, there are variations in the local geometry of the bases. The propeller twist and buckle are the dihedral angles formed between the individual bases when viewed along the long and short axis respectively. The d(CGCICG) structure, like the two d(CGCGCG) structures, follows the same trend in exhibiting the characterstic negative propeller twist for the CpG steps and positive twist for GpC steps. The buckle angles, which are another measure of the noncoplanarity, are significantly different for the individual base-pairs but follow a similar trend in all three structures. However, values for the individual base pairs in d(CGCICG)structure are larger in magnitude than in the d(CGCGCG) structures. The central CpI and IpC base pairs show large negative buckle and positive buckle angles respectively indicating these base-pairs are bent toward one another like cupped hands. These high values may be a consequence of optimizing the stacking interactions with

the sugar phosphate backbone of other symmetry related molecules, and/or due to the different crystal lattice forces. The x-displacement values between individual pairs of base pairs follow the same trend for all three structures, although the extent of x-displacement for individual base-pairs varies significantly for all three structures. The higher x-displacement for the base-pairs in d(CGCICG) DNA indicates the base-pairs are moved away from the helix axis toward the major groove ie. the helical axis lies within the minor groove instead of passing through the O2 oxygens of cytosines. The difference between average x-displacement for d(CGCICG) and the mixed magnesium/spermine form d(CGCGCG) is 0.81 Å and the corresponding difference between d(CGCICG) and magnesium form d(CGCGCG) is 0.60 Å. There is a significant variation in the x-displacement values between individual base-pairs, and the net effect of these uneven offsets between individual base-pairs results in different overlaps between stacked base-pairs. The tip angle, which defines the rotation about the long axis of the base-pair perpendicular to the helix axis shows substantial variation for all three structures. The terminal base-pairs in d(CGCICG) DNA are tipped the most and are most probably influenced by the stacking of these base-pairs against the sugar-phosphate backbone of symmetry-related molecules. The tip angles in all three structures however, follow the same trend ie. alternation of high values for pyrimidine-purine steps and low value for the purine-pyrimidine steps. Similarly the roll angles also show substantial variation in the three structures. The roll angles for the purine-pyrimidine (G:C and I:C base-pair) steps in d(CGCICG) DNA are larger in magnitude compared to d(CGCGCG) DNA structures. The most interesting feature is that the roll angles for the individual base-pair steps in all three Z-DNA structures show an alternation with low values for purine-pyrimidine steps and high values for pyrimidine-purine steps. These values are in contrast to the predictions made by Calladine (24) for an alternating sequence. The average inclination angle for all three structures are similar. However, the mixed magnesium/spermine form d(CGCGCG) shows relatively high inclination angles for individual base-pairs.

Backbone conformation

Table III compares the sugar-phosphate backbone, and glycosidic torsion angles of d(CGCICG) DNA with the mixed magnesium/spermine and magnesium forms of d(CGCGCG) DNA. Differences among the torsion angles of all three structures are generally small. However, certain torsion angles of the d(CGCICG) structure are distinctly different from the parent d(CGCGCG) structures. The average pseudorotation phase angle of deoxyribose rings of the cytidines in d(CGCGCG) structures is 154°, typical of a C2′-*endo* pucker, while the corresponding average pseudorotation angle for the guanosines is 25°, which is typical of a C3′-*endo* conformation. However, in d(CGCICG) DNA the pseudorotation angle for the cytidine C5 is 132° which is typical of C1′-*exo* conformation, while the inosine residues I4 and I10 average 58°, which suggests they are in C4′-*exo* conformation. The switch of the sugar pucker of inosine residues from C3′- *endo* to C4′-*exo* not only affects the local base parameters, but also perturbs the neighboring base pairs. It is very likely that this distortion is brought about by crystal packing. The average glycosidic torsion angle χ for the pyrimidines in d(CGCICG) DNA is 204°, which is slightly different from the 208° for the two d(CGCGCG) DNA structures, but still consistent with the *anti* conformation. The purines adopt a *syn* conformation: average χ angle for d(CGCICG) DNA is 63° compared to 70° and 65° for the mixed magnesium/spermine and magnesium forms of d(CGCGCG) DNA respectively.

The P...P distances range from 5.74 Å to 6.7 Å, with an average of 6.2 Å. The average P...P values for mixed magnesium/spermine form d(CGCGCG) and magnesium d(CGCGCG) DNA are 6.3 and 6.35 Å respectively. The compression of the helical axis and a narrower groove for d(CGCICG) DNA is apparent from the phosphate-phosphate distances. The average intrastrand phosphate-phosphate distances in the inosine form is 9.14 Å, which is shorter than 9.87 Å and 9.90 Å for the mixed magnesium/spermine and magnesium form d(CGCGCG) DNA respectively. Similarly, the average interstrand or the distances between phosphates of different strands for the inosine-form d(CGCICG) DNA is 9.2 Å compared to 9.82 Å and 9.81 Å for the mixed magnesium/spermine and magnesium form d(CGCGCG) DNA respectively, and thus resulting in a narrower minor groove for the inosine-form d(CGCICG) DNA. The phosphate groups of GpC are observed in two conformations in Z-DNA (25). These two conformations have been called Z_I and Z_{II} respectively; conformation Z_I is *synclical (−),gauche(−)-trans* for the phosphodiester conformation,while Z_{II} is *synclical(+), gauche(+)-trans*. The difference between Z_I and Z_{II} is most pronounced in torsion angles ϵ and ζ and to a lesser extent in $\alpha$and β angles. The backbone torsion angles of the two d(CGCGCG) DNA structures around the phosphate of residue C5 are similar, but differ from the d(CGCICG) structure. The phosphates linking I4 and C5 in strand 1, and I10 and C11 in strand 2 are rotated toward the center of the groove in the Z_I position. For this conformation, ζ angles of I4 and I10 are in $-sc$ conformation. A similar phosphate conformation was seen for the pure-spermine d(CGCGCG) DNA. The phosphate linking G10 and C11 in d(CGCGCG) DNA is in a similar conformation but the phosphate linking G4 and C5 in the two d(CGCGCG) structures is rotated downward away from the groove in the Z_{II} position. The corresponding ζ angle of G4 is in $+sc$ conformation.

The crystal packing for the d(CGCICG) structure shows similar interactions to those observed with other d(CGCGCG) structures. However, analysis revealed several interesting features (1) The unit cell volume for d(CGCICG) DNA is 24220 Å^2, lower than the values of 25148 Å^2, 25036 Å^2, and 24436 Å^2 observed for the mixed magnesium/spermine, magnesium and pure spermine forms respectively. (2) Analysis of the unit cell dimensions of d(CGCICG) DNA shows that the differences are greatest in the *c* cell length (43.3 Å), which in the mixed magnesium/spermine and magnesium forms average 44.76 Å and 44.38 Å respectively. These observations suggest that the decrease in the cell volume is a result of compression along the helical axis (approximately the same as the *z* axes. A similar observation was made for the pure spermine form of d(CGCGCG) DNA (26), Z-RNA and (araC)-Z-DNA (27). It is also interesting to note that although the duplex is rotated upon shifting from the mixed magnesium/spermine lattice to the inosine DNA lattice, the stacking interactions between adjacent duplexes in the lattice are not altered significantly. In all of these lattices the duplexes stack in an anti-parallel (5′−3′ and 3′−5′) manner, forming infinite helices along the *z* direction. There are two intermolecular contacts between the molecules, and both involve adjacent duplexes: The O1P atom of residue G(4) and O4′ of G(8) of one duplex forms a hydrogen bond with O3′ of G(6) and O4′ of G(10) of an adjacent duplex respectively. As reported by Egli et al.,

(26), the 3′ termini of d(CGCGCG) DNA structures show pronounced polymorphism. In the pure-spermine form, the O3′ of G(6) forms a direct hydrogen bond to O1P of G(2) of neighboring duplex, while in the magnesium form, the O3′ atom of residue G(6) forms a hydrogen bond with O2P of C(9) of the adjacent duplex. In contrast, the O3′ of G(6) in mixed magnesium/spermine form interacts indirectly with O2P of C(9) of an adjacent duplex via an intermediate water molecule. The O3′ atom of residue G(12) forms a direct hydrogen bond with O2P of G(2) from a neigboring duplex in both the mixed magnesium/spermine and magnesium forms. While in the inosine form, hydrogen bonding between O3′ of G(12) and O2P of G(2) is via an intermediate water molecule. In the pure spermine form the O3′ atom of G(12) does not interact with neighboring duplexes. In addition we observe C-H..O hydrogen bonds involving C8 of G(8) with O2P of residue C(9) of an adjacent duplex. Similar observations were seen in the pure spermine form of d(CGCGCG) DNA (26).

Thermal motions and solvent structure

The isotropic temperature factors (*B* factors) reflect the mobility of the atoms within the crystal. The average thermal parameters of the bases (15.4 Å^2) are lower than sugars (24 Å^2) or phosphate groups (33 Å^2), reflecting lesser mobility for the bases as generally observed for crystal structures of oligonucleotides. In the course of refining the structure to 1.7 Å resolution, 70 water molecules were located in the assymetric unit of the d(CGCICG) DNA structure. The solvent thermal parameters range from 23 Å^2 to 55 Å^2. Several well-defined water molecules were observed along the helical axis, including the minor groove of the d(C:I) base pair, consistent with observations for all of the Z-DNA structures studied so far. These waters form a continuous spline along the minor groove in the case of the d(CGCGCG) structure, which thought to be important for stabilizing Z-DNA in this structure. All phosphate anionic oxygen atoms accept at least one hydrogen bond from a water molecule. The phosphate oxygens of residues 3 and 12 accept five and six hydrogen bonds respectively, whereas the phosphate groups of residues 4, 5, 6 accept four hydrogen bonds. Forty five of the solvent molecules are involved in inter-molecular contacts, of which sixteen contact symmetry related molecules. In the mixed magnesium/spermine form there are eleven such contacts, while in the magnesium form there are thirteen such water molecules forming inter-molecular contacts. There are twelve water molecules with two or more contacts to the same DNA duplex, and the positions of five of these water molecules are very similar to either water positions in both the magnesium or mixed magnesium/spermine form d(CGCGCG) DNA. Comparison with the pure-spermine form show a higher conservation of the water positions, with eight of these water molecules forming similar contacts with the DNA duplex.

CONCLUSIONS

The present analysis establishes that the hexamer duplex d(CGCICG) like the parent d(CGCGCG), forms a Z-type structure under the crystallization conditions employed. A comparison of the inosine-containing d(CGCICG) DNA molecule with the mixed magnesium/spermine and magnesium form Z-DNA structures reveals minor differences in the helical parameters and backbone. The crystal structure of d(CGCICG) shows that inosine can replace guanine and still retain the conformation of the parent DNA. The most striking aspect of the Z-DNA molecules is the virtual identity of the corresponding base-steps in all three helices, even though they were crystallized under slightly different conditions, and their crystal structures were refined independently. Although all three structures share this fundamental similarity in that the base-pairs stack in an identical manner, the crystal packing interactions, thermal parameters, and base stacking interactions are quite different in all three crystal forms. Many of these differences in the three crystal forms can be explained by the crystal packing interactions, which are responsible for distortions of the duplexes at different locations (28). The most noteworthy features of the inosine-containing DNA structure as result of such distortions are (1) the helix shows a rotation of 14.79° and a translation of 16.8 Å along the helical axis compared to the mixed magnesium/spermine form, (2) all phosphates have the Z_I conformation, and (3) narrower minor groove and compression along the helical axis. and (4) the sugar puckering for the inosines is of the C4′-*exo* type In addition, the substitution of guanosine by inosine appears to have resulted in Watson−Crick type base-pairing between inosine and cytidine with a potential bifurcated hydrogen bond between inosine N1 and cytidine N3 (2.9 Å) and O2 (3.3−3.Å). In conclusion we have observed that d(CGCICG) DNA crystallizes in an orthorhombic $P2_12_12_1$ lattice that more closely resembles the pure-spermine form of d(CGCGCG) DNA rather than the mixed magnesium/spermine or magnesium forms of d(CGCGCG) Z-DNA structures.

ACKNOWLEDGEMENTS

We thank Dr Eric Wickstrom for helpful suggestions, Dr R.E.Dickerson for providing us with the NEWHEL91 program, Marian Powers for synthesizing oligonucleotide, and the Advanced Scientific Computing laboratory at NCI-FCRDC, Frederick, MD for the Cray time to run XPLOR.

REFERENCES

1. Martin, F. H., Castro, M. M., Aboul-ela, F., and Tinoco, Jr., I. (1985) *Nucleic Acids Res* **13** 8927−8938
2. Felsenfeld, G. and Miles, H. T. (1967) *Ann. Rev. Biochemistry.* **31**, 407−448.
3. Hall, K., Cruz, P., Chamberlin, M. J. (1985) *Arch. Biochem. Biphys.* **236**, 47−51.
4. Takahashi, Y., Kato,K., Hayashizaki, Y., Wakabayashi, T., Ohtsuka, E. Matsuki,S., Ikehara, M., and Matsubara, K. (1985) *Proc. Natl. Acad. Sci.* **82**, 1931−1935.
5. Corfield, P. W. R., Hunter, W. N., Brown, T., Robinson, P. and Kennard, O. (1987) *Nucleic Acids Res.* 15, 7935−7949.
6. Cruse, W. B. T., Aymani, J., Kennard, O., Brown, T., Jack, A. G. C. and Leonard, G. A. (1989) *Nucleic Acids Res.* **17**, 55−72.
7. Kumar, V. D., Harrison, R. W., Andrews, L. and Weber, I. T. (1992) *Biochemistry.* **31**, 1541−1550.
8. Xuan, J-C. and Weber, I. T. (1992) *Nucleic Acids Res.* **20**, 5457−5464.
9. Pohl, F. M. and Jovin, T. M. (1972) *J. Mol. Biol.* **67**, 357−396.
10. Wang, A. H. -J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. and Rich, A. (1979) *Nature.* **282**, 680−686.
11. Drew, H., Takano, T., Tanaka, S., Itakura, K. and Dickerson, R. E. (1980) *Nature.* 286, 567−573.
12. Leslie, A. G. W., Arnott, S. Chandrasekaran, R. and Ratliff, R. L. (1980) *J. Mol. Biol.* 143, 49−72.
13. Mitsui, Y., Langridge, R., Shortle, B. E., Cantor, C. R., Grant, R. C., Kodama, M. and Wells, R. D. (1970) *Nature.* **228**, 1166−1169.
14. Vorlickova, M. and Sagi, J. (1991 *Nucleic Acids Res.* 19, 2343−2347.
15. Gessner, R. V., Frederick, C. A., Quigley, G. J., Rich, A. and Wang, A. H. -J. (1989) *J. Biol. Chem.* **264**, 7921−7935.
16. Higashi, T. (1990) *PROCESS: A program for indexing and processing R-AXIS II imaging plate data,* Rigaku Corporation, Japan.

17. Brunger, A. T., Kuriyan, J., and Karplus, M. (1987) *Science.* **235**, 458–460.
18. Lattman, E. E. (1985) *Methods Enzymol.* **115**, 55–77.
19. Brunger, A. T. (1990) *Acta. Cryst.* A. **A46**, 46–57.
20. Jones, T. A. (1985) *Methods Enzymol.* **115**, 157–171.
21. Hendrickson, W. A., and Konnert, J. H. (1981) *in Biomolecular Structure, Conformation, Function and Evolution, Vol. 1.* (Srinivasan, R. Ed.), pp 43–57, Pergammon, Oxford, U.K.
22. Westhoff, E., Dumas, P., and Moras, D. (1985) *J. Mol. Biol.* **184**, 119–145.
23. Kennard,O. (1988) *in Structure and Expression, Vol. 2.* (Sarma,R. H. and Sarma, M. H., Eds.), pp 1–25, Academic Press, New York.
24. Calladine, C. R. (1982) *J. Mol. Biol.* **161**, 343–352.
25. Wang, A. H. -J., Quigley,G. J., Kolpak,F. J., van der Marel, G., van Boom, J. H., and Rich, A. (1981) *Science.* **211**, 171–176.
26. Egli, M., Williams, L. D., Gao, Q. and Rich, A. (1991) *Biochemistry.* **30**, 11388–11402.
27. Teng, M. K., Liaw, Y. C., van der Marel, G., van Boom, J. H., and Wang, A. H.-J. (1989) *Biochemistry.* **28**, 4923–4928.
28. Jain, S. and Sundarlingam, M. (1989) *J. Biol. Chem.* **264**, 12780–12784.